The MAQC Project:

MicroArray Quality Control

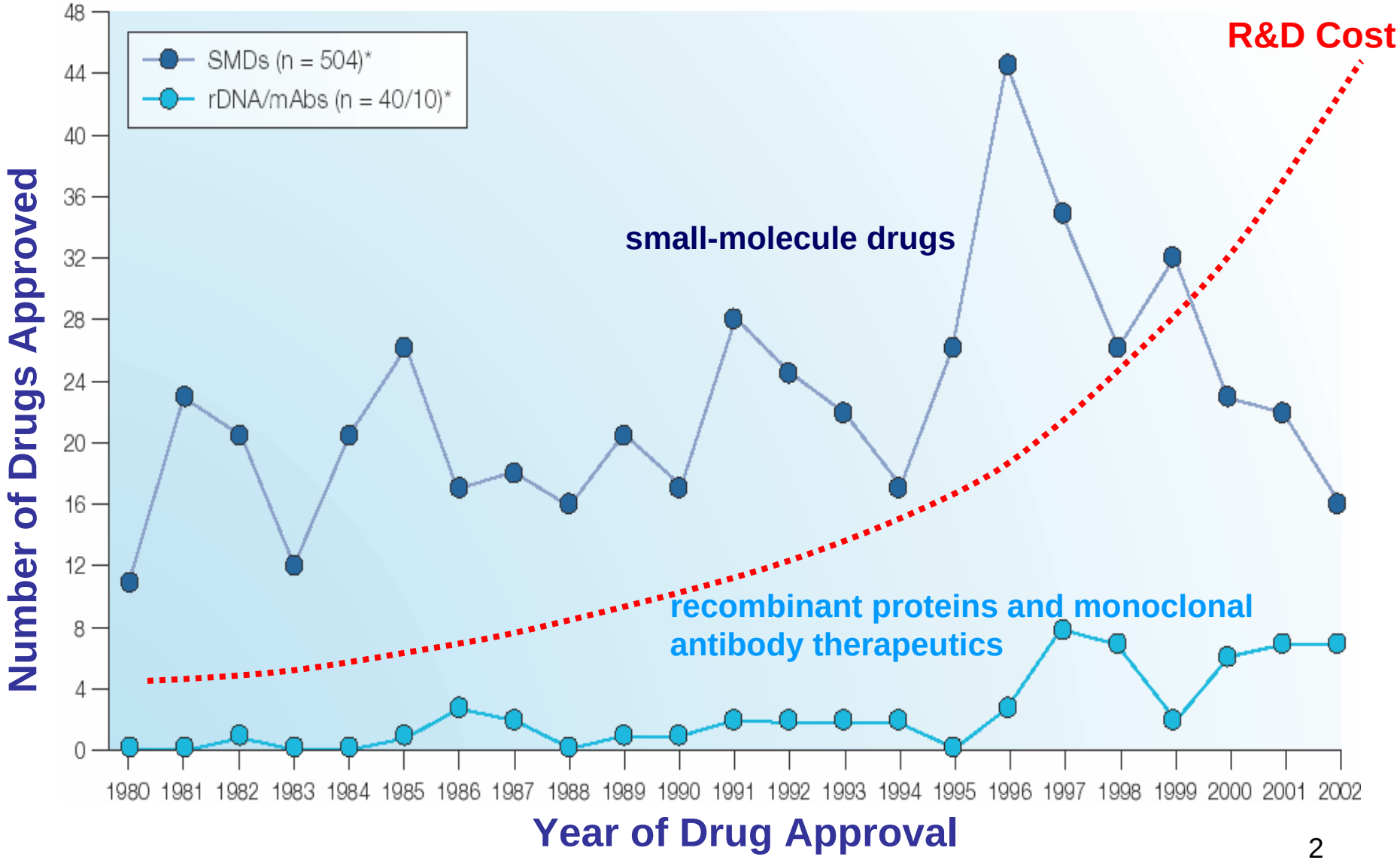
Leming Shi	FDA/NCTR
Shawn C. Baker	Illumina
Kathy Y. Lee	Applied Biosystems
Yuling Luo	Genospectra (Panomics)
James C. Willey	Ohio Medical Univ./Gene Express
Paul K. Wolber	Agilent

http://edkb.fda.gov/MAQC/

CHI's 2nd Annual QPCR Meeting
San Diego, CA, March 20, 2006

Views expressed in this presentation are those of the presenter and not necessarily those of the U.S. FDA.

Decline of the Number of Drugs Approved in the USA



Modified from Reichert JM, *Nature Review Drug Discovery* 2003

U.S. Food and Drug Administration

Department of Health and Human Services

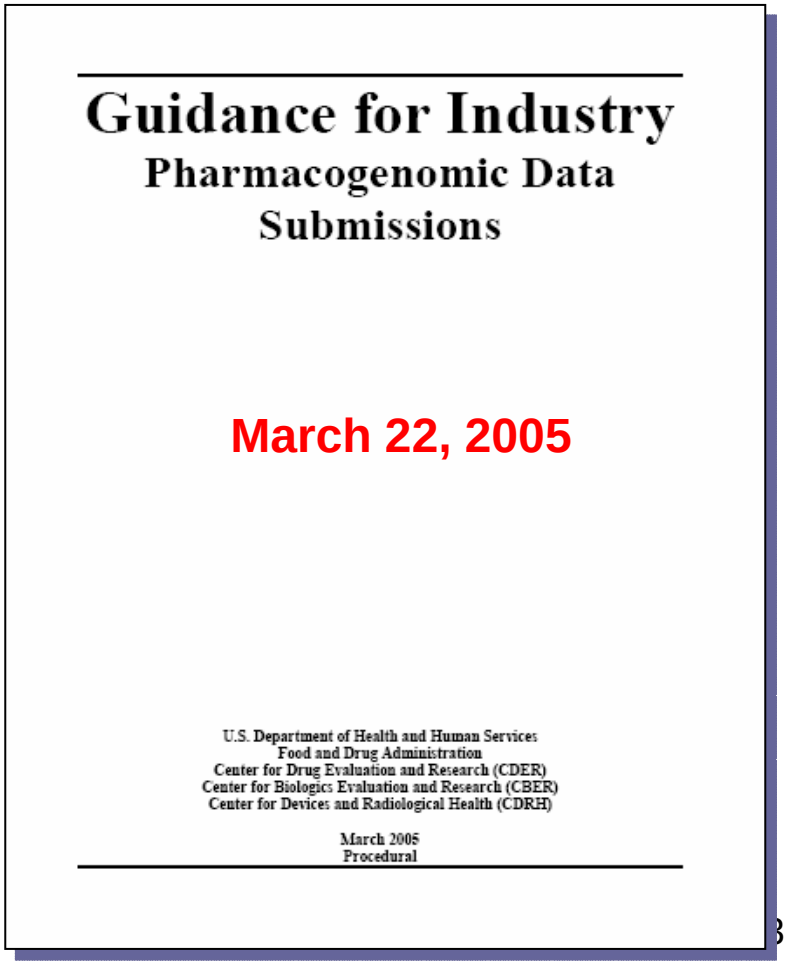
The Critical Path to New Medical Products

http://www.fda.gov/oc/initiatives/criticalpath/
http://www.fda.gov/cder/genomics/

76 Opportunities

1. Biomarker Qualification
2. **Standards for Microarray** and Proteomics-Based Identification of Biomarkers

……



FDA pharmacogenomics guidance sends clear message to industry

Pharmacogenomic information will be an essential element in drug submissions.

Mark Ratner

After much anticipation, the US Food and Drug Administration delivered its final guidance document on Pharmacogenomic Data Submissions on 22 March.

Companies have eagerly awaited the

BIOMARKER DEFINITIONS

Valid biomarker. A biomarker that is measured in an analytica
established performance characteristics.
Known valid biomarker. Widespread agreement in the medica
the physiological, toxicological, pharmacological and/or clinic
Probable valid biomarker. A scientific framework or body of ev

Three pages later …

"FDA doesn't spend its resources, which are in constant stress in a variety of directions, to work hard and issue this kind of document unless they see it as the future," suggests Samuel Broder, Chief Medical Officer at Celera Genomics.

"FDA has now made it clear that as the technology base grows, it will move toward stricter regulatory rules," says Drews. The decision by FDA to publish "is a definitive step in favour of the progressive forces in industry and biotech."

NATURE REVIEWS | DRUG DISCOVERY 362 | MAY 2005 | VOLUME 4

An array of problems

Despite the huge amount of published microarray data in cancer, little is being converted into clinical practice. Validating initial data is proving to be a key challenge, reports SIMON FRANTZ.

Study reference	Cancer type	Clinical endpoint	Sample size	Number of events (%)	Number of channels (type)	Number of genes after filtration*
2	Non-Hodgkin lymphoma	Survival	240	138 (58%)	2 (Lymphochip)	6693
3	Acute lymphocytic leukaemia	Relapse-free survival	233	32 (14%)	1 (Affymetrix)	12 236
4	Breast cancer	5-year metastasis-free survival	97	46 (47%)	2 (Agilent)	4948
5	Lung adenocarcinoma	Survival	86	24 (28%)	1 (Affymetrix)	6532
6,7	Lung adenocarcinoma	4-year survival	62†	31 (50%)	1 (Affymetrix)	5403
8	Medulloblastoma	Survival	60	21 (35%)	1 (Affymetrix)	6778
9	Hepatocellular carcinoma	1-year recurrence-free survival	60	20 (33%)	1 (Affymetrix)	4861

*For the data of van 't Veer and colleagues,[4] the same filter was used as in the original publication. For other studies, genes with little variation in expression were excluded. †Only patients with clinical follow-up of at least 4 years after surgical resection were analysed.[7]

Table: Description of eligible studies ordered by sample size

2 Rosenwald A, et al. *N Engl J Med* 2002; **346:** 1937–47.
3 Yeoh EJ, et al. *Cancer Cell* 2002; **1:** 133–43.
4 van't Veer LJ, et al. *Nature* 2002; **415:** 530–36.
5 Beer DG, et al. *Nat Med* 2002; **8:** 816–24.
6 Bhattacharjee A, et al. *PNAS* 2001; **98:** 13790–95.
7 Ramaswamy S, et al. *Nat Genet* 2003; **33:** 49–54.
8 Pomeroy SL, et al. *Nature* 2002**; 415:** 436–42.
9 Iizuka N, et al. *Lancet* 2003; **361:** 923–29.

"Five of the seven studies did not classify patients better than chance."

"We found that the list of genes that are most differentially expressed in the populations studied varied greatly." [Catherine Hill]

1. Michiels, S. *et al. Lancet* **365**, 488–492 (2005); 2. Ioannidis, J. P. A. *Lancet* **365**, 454–455 (2005).

SPECIAL SECTION

GENES IN ACTION

NEWS

Getting the Noise Out of Gene Arrays

Thousands of papers have reported results obtained using gene arrays, which track the activity of multiple genes simultaneously. But are these results reproducible?

he gathered on kidney tumor cells, the less significant it seemed.

But those who have persevered with gene expression arrays attribute such problems to early growing pains. They claim

E. Marshall, *Science* 306, 630 (Oct 22, 2004).

"**Little overlap**."

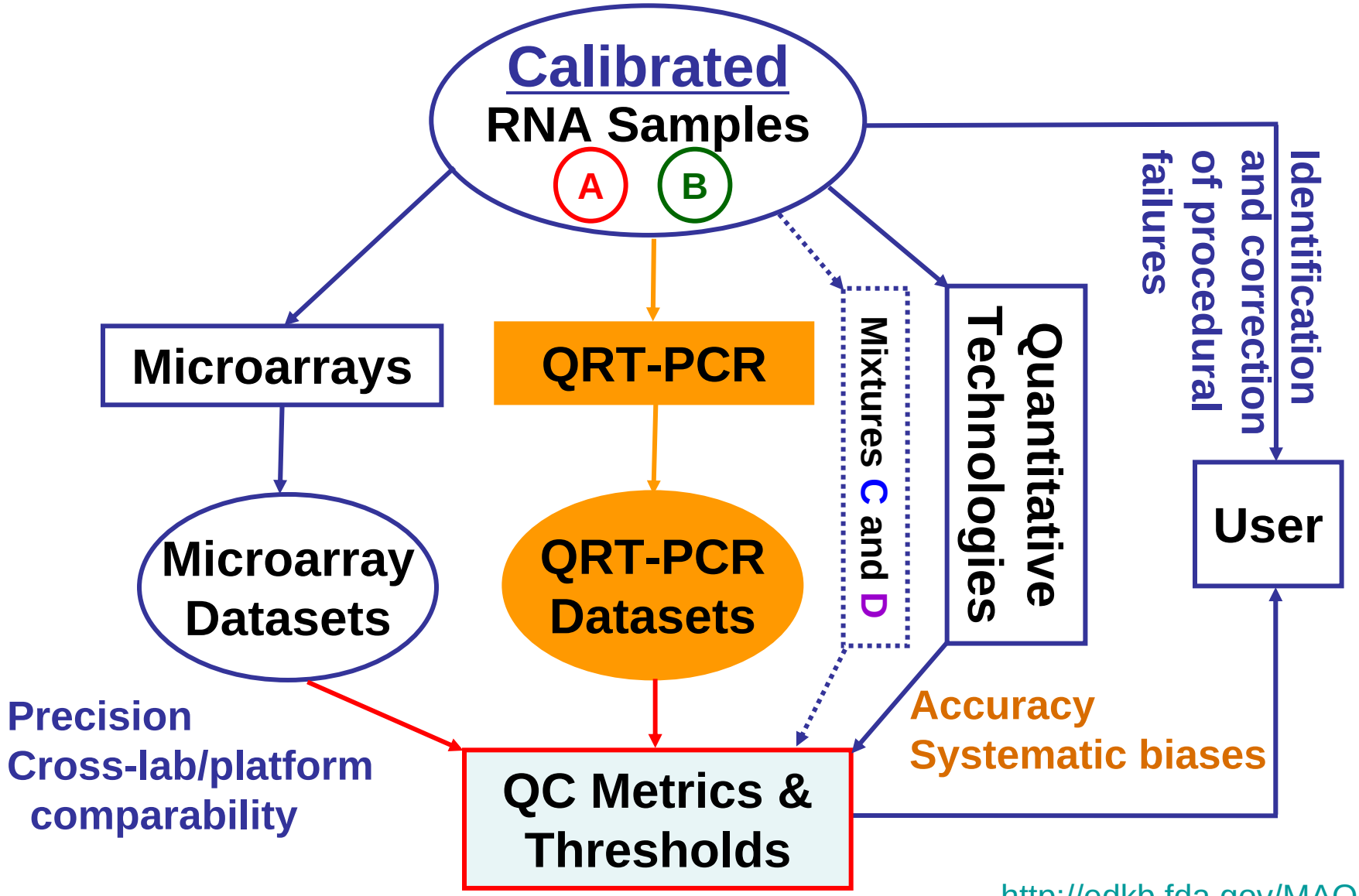
"… suggesting the need for establishing industrial **manufacturing standards**, and further independent and thorough validation of the technology."

P.K. Tan *et al.*, *Nucleic Acids Res* 31, 5676 (Oct 1, 2003).

Two Challenges Towards Reproducibility

- **Ensuring experimental data quality of individual laboratories:**
  - **Assessing the best achievable technical performance of microarray platforms (QC metrics and thresholds)**
- **Reaching consensus on data analysis:**
  - **Assessing the advantages and disadvantages of various data analysis methods**

The MAQC Project: MicroArray Quality Control



http://edkb.fda.gov/MAQC/

Evaluation of data analysis methods

MAQC

>1,000 arrays

Pilot-I: RNA Samples

- 1st face-to-face meeting on Feb. 11, 2005 at FDA/NCTR
- Selection of two RNA samples from four candidates
- Five replicate microarrays for each RNA
- Four microarray platforms (AFX/AGL/GEH/ILM)
- **160** microarrays from seven test sites
- 2nd face-to-face meeting on May 2-3, 2005 at FDA/CDER

A: Stratagene UHRR; **B**: Ambion Brain RNA

Completed April 2005

Pilot-II: Tissue Titration

- Selection of two of the 13 mixtures of **A** and **B**
- Three to five replicate microarrays for each mixture
- Four microarray platforms (AFX/AGL/GEH/ ILM)
- Two alternative platforms (TAQ/QGN)
- **200** microarrays from four platform providers
- TAQ/QGN validation data for ten tissue-specific genes

Completed July 2005

C: 75%A+25%B; **D**: 25%A+75%B

Main Study: Reference Datasets

- Four RNA samples (**A**, **B**, **C**, and **D**)
- Detailed QC data for total RNA and targets
- Five replicate microarrays for each RNA
- Seven microarray platforms (ABI/AFX/AGL/EPP/GEH/ILM/NCI)
- Three to six test sites for each microarray platform
- **534** microarrays from 24 test sites
- Three alternative technology platforms (TAQ/QGN/GEX)
- **1000** genes by TAQ; **245** genes by QGN; **207** genes by GEX

200 addl. arrays

Completed Oct. 2005

Microarray Platforms:

- **ABI**: Applied Biosystems
- **AFX**: Affymetrix
- **AGL**: Agilent
- **EPP**: Eppendorf
- **GEH**: GE Healthcare
- **ILM**: Illumina
- **NCI**: NCI_Operon custom oligoarray

Alternative Technology Platforms:

- **GEX**: StaRT-PCR from Gene Express
- **QGN**: QuantiGene from Genospectra
- **TAQ**: TaqMan® from Applied Biosystems

Data Analysis

- ~40 organizations are analyzing the datasets
- QC metrics and thresholds
- Precision and cross-lab/platform comparison
- Sequence-based cross-platform mapping
- Normalization and gene selection methods
- Validation of microarray results
- Titration datasets for assessing accuracy
- Performance of spike-in controls
- Modeling cross-hybridization
- One-color versus two-color designs
- Informatics tools
- Public deposition
- MAQC-3 in Palo Alto, CA, Dec. 1-2, 2005
- MAQC-4 in Boston, MA, Feb. 3-4, 2006

MAQC Publications (Sept. 2006)
MAQC Guidance (2006-2007?)

Eleven (11) Research Manuscripts Are Being Prepared for Submission to *Nature Biotechnology* by May-31-2006

1. **MAQC Main Paper**
2. **Probe Sequence Mapping (RefSeq)**
3. **Probe Sequence Mapping (AceView)**
4. **The Stability/Reproducibility of Signature Gene Lists**
5. **Validation of Microarray Results**
6. **Titration and Relative Accuracy**
7. **Modeling Technical Variation**
8. **Reproducibility Analysis**
9. **One-color versus Two-color Designs**
10. **Spike-in Controls for Quality Assessment**
11. **Validation with Rat Toxicogenomics Data Sets**

Plus: Editorial, Foreword, Commentary, Perspectives (FDA and EPA)

MAQC Participating Organizations

Government Agencies (4):
FDA (six Centers)
EPA
NIH (NCBI and NCI)
NIST

Platform Providers (10):
Affymetrix (AFX)
Agilent (AGL)
Applied Biosystems (ABI)
Eppendorf (EPP)
GE Healthcare (GEH)
Illumina (ILM)
National Cancer Institute (NCI)
Applied Biosystems' TaqMan (TAQ)
Gene Express' StaRT-PCR (GEX)
Genospectra's QuantiGene (QGN)

RNA Sample Providers (3):
Ambion
Clontech
Stratagene

http://edkb.fda.gov/MAQC/

Test Sites (>30):
ABI_1: Applied Biosystems
ABI_2: EPA
ABI_3: Vanderbilt University
AFX_1: Affymetrix
AFX_2: FDA/CDER
AFX_3: Ambion
AFX_4: EPA
AFX_5: Novartis
AFX_6: UCLA/Cedars-Sinai
AGL_1: Agilent
AGL_2: FDA/NCTR
AGL_3: Icoria
EPP_1: Eppendorf
EPP_2: MD Anderson
EPP_3: CSHL
GEH_1: GE Healthcare
GEH_2: UMass Boston
GEH_3: GenUs BioSystems
ILM_1: Illumina
ILM_2: UT Southwestern
ILM_3: Burnham Institute
NCI_1: NIH/NCI
NCI_2: FDA/NCTR
NCI_3: FDA/CBER
TAQ_1: Applied Biosystems
QGN_1: Genospectra
GEX_1: Gene Express

Data Analysis Sites (11):

Biogen Idec
Expression Analysis
FDA/NCTR
Harvard University
NIH/NCBI
NIST
SAS
Stanford University
UIUC
UMass Boston
ViaLogy

MAQC Mailing List:
Over **200** people from >**70** organizations

CapitalBio
Harvard
Operon
TeleChem
Wake Forrest Univ.
Yale Univ.

ERCC